Visualising vowel harmony decay in Old Norse manuscripts

Jade J. Sandstedt
jsandstedt@gmail.com
jsandstedt.hcommons.org

10. December 2019

Contents

1 Introduction .................................................. 2
   1.1 Vowel harmony basics ................................... 2
   1.2 Vowel harmony decay .................................... 3
   1.3 Sources of evidence ...................................... 3

2 Corpus study ................................................. 4
   2.1 Old Norwegian vowel harmony ............................ 4
   2.2 Methods and corpus .................................... 7
   2.3 Sanity check ........................................... 10
       2.3.1 Vowel and word length frequencies .............. 10

3 Visualising vowel harmony decay .......................... 12
   3.1 *PhonMatrix* visualisations ............................ 12
   3.2 Distinguishing decay stages ........................... 15

4 Conclusions ................................................ 16
Outline

Topic:

• Tracking vowel harmony changes in written corpora

Methods and solutions:

• rich linguistic annotations
• novel visualisations
• working diagnostics

I Introduction

1.1 Vowel harmony basics

Very generally defined, vowel harmony is a process in which vowels in a word show systematic correspondence for some feature.

• an example of labial or rounding harmony is provided in (i)

(i) Rounding harmony in Yakut (Siberian–Turkic; Krueger 1962: pp. 46–53)
  a. kel-el-ler ‘come’-3.PRES.-PL.
  b. kor-ol-lor ‘see’-3.PRES.-PL.
  c. kele-\-\-in ‘come’-2.SG.
  d. dojo-\-yn ‘grow quiet’-2.SG.

Vowel harmony typology

Any segmental feature may serve as the basis for a harmony system

Chewa (Bantu) height harmony (Downing & Mtenje 2017)
  [+high] ph\-\-k-il ‘cook’-APPL.
  [−high] ts\-\-k-el ‘close’-APPL.

Finnish (Finno–Ugric) backness harmony (Ringen 1975)
  [+back] p\-\-ou\-ta-na ‘fine weather’-ESS.
  [−back] p\-\-ytn\-\-n\-\-a ‘table’-ESS.

Yoruba (Atlantic–Congo) tongue root harmony (Ola Orie 2001, 2003)
  [+ATR] o\-g\-\-de ‘incantations’
  [−ATR] go\-g\-\-de ‘banana, plantain’
Prevalence and motivations for harmony

Harmony systems are

• articulatorily and perceptually motivated,
  – eases articulation, makes sequences more predictable, enhances perceptually weak cues, etc. (Suomi 1983, Gallagher 2010, Walker 2005)

• easy to learn and acquired early
  – few to no harmony violations by ca. 2;6 years (MacWhinney 1978, Leiwo, Kulju & Aoyama 2002, Altan 2007)

• cross-linguistically very common and diachronically robust,
  – e.g. millenia old backness harmony in Turkic languages (Harrison, Dras & Kapicioğlu 2006)

1.2 Vowel harmony decay

Harmony doesn’t last forever

Despite the stability of harmony systems, diachronic and/or cross-dialectal correspondences with historical and existing harmony languages show that harmony systems do decay.

- e.g. Turkish vs. Uzbek (Turkic; Csató & Johanson 1998; Sjoberg 1963).

(2) Turkic backness harmony lost in Uzbek

<table>
<thead>
<tr>
<th>BACK</th>
<th>dost-lar</th>
<th>‘friend’-pl.</th>
<th>dost-lar</th>
<th>‘friend’-pl.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kul-lar</td>
<td></td>
<td>‘slave’-pl.</td>
<td>Kul-lar</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FRONT</th>
<th>et-lar</th>
<th>‘meat’-pl.</th>
<th>et-lar</th>
<th>‘meat’-pl.</th>
</tr>
</thead>
<tbody>
<tr>
<td>*et-lar</td>
<td></td>
<td></td>
<td>*et-lar</td>
<td></td>
</tr>
<tr>
<td>dis-lar</td>
<td>*dis-lar</td>
<td>‘tooth’-pl.</td>
<td>dis-lar</td>
<td></td>
</tr>
<tr>
<td>tish-lar</td>
<td>*tish-lar</td>
<td>‘tooth’-pl.</td>
<td>tish-lar</td>
<td></td>
</tr>
</tbody>
</table>

(a) Turkish – [-lar] / [-ler]  (b) Uzbek – [-lar]

Question: If harmony is so natural and beneficial, what motivates harmony decay and how do harmony processes die?

1.3 Sources of evidence

We know currently little about the causes and nature of harmony decay

• no historical record has been shown to demonstrate harmony decay in progress.
Currently, we can examine harmony decay using:

- comparisons between harmonic/non-harmonic dialects
  - Crimean Tatar (Turkic; Kavitskaya 2013)

- diachronic comparisons before and following harmony decay
  - Kazakh (Turkic; McCollum 2015); Itelmen (Chukotko-Kamchatkan; Bobaljik 2018)

- agent-based computational modelling of potential trajectories of vowel harmony evolution/decay
  - e.g. Harrison, Dras & Kapicioglu (2006); Mailhot (2010)

Suspected causes of decay:

- changes in vowel inventories (mergers/splits),
- emergence of disharmonic morphemes,
- language contact (i.e. via the influx of disharmonic foreign loanwords)

<table>
<thead>
<tr>
<th>Problem: We lack empirical evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>☞ crucial missing link in the typological record: the transition from a harmonic to non-harmonic language.</td>
</tr>
<tr>
<td>☞ unclear how and why these factors might converge on the loss of harmony</td>
</tr>
</tbody>
</table>

2 Corpus study

2.1 Old Norwegian vowel harmony

Old Norwegian (c. 1200–1350) displays a form of vowel height harmony (j)

- resulting in [-i]/[-e] and [-u]/[-o] suffixal alternations

<table>
<thead>
<tr>
<th>(j) Height harmony in Old Norwegian (Sandstedt 2017, 2018)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HIGH</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>NON-HIGH</strong></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

• cf. non-harmonic Old Icelandic **ljós-i** and **segl-i**
Old Norwegian vowel harmony typology
Old Norwegian vowel harmony is most similar to Bantu height harmony systems in central and southern Africa

- e.g. Mbunda (K.15; aka Chimbunda, Kimbunda, or Mbuunda), spoken in Angola and Zambia (Gowlett 1970).

(4) Mbunda height harmony on APPL. /-il-/  
HIGH lum-il- ‘cultivate’ tung-il- ‘build’  
MID nen-el- ‘bring’ oc-el- ‘roast’  
LOW kwat-el- ‘hold’

Old Norse corpus
There is a diverse Old Norwegian corpus from the harmony period

- including sizeable manuscripts (Fig. 2)  
- dated charters (Fig. 3)  
- and original runic inscriptions (Fig. 4)

Figure 2: Old Norwegian Homily Book (AM 619 4to; c 1200)
Figure 3: A charter from King Magnus VII – Niðaróss (Trondheim), 29. Jan. 1333

Figure 4: Runestone from Kingittorsuaq (Greenland) with height harmony

5 Early Old Norse height harmony lost in Icelandic

<table>
<thead>
<tr>
<th></th>
<th>High</th>
<th>Non-/</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>hús-um</td>
<td>'house'-DAT.PL.</td>
<td>ljós-om</td>
<td>'light'-DAT.PL.</td>
</tr>
<tr>
<td>skip-um</td>
<td>'ship'-DAT.PL.</td>
<td>segl-um</td>
<td>'sail'-DAT.PL.</td>
</tr>
<tr>
<td>hús-um</td>
<td>'house'-DAT.PL.</td>
<td>ljós-um</td>
<td>'light'-DAT.PL.</td>
</tr>
<tr>
<td>skip-um</td>
<td>'ship'-DAT.PL.</td>
<td>segl-um</td>
<td>'sail'-DAT.PL.</td>
</tr>
</tbody>
</table>

(a) Old Norwegian – [-um] / [-om]  
(b) (Old) Icelandic – [-um]

It is currently conjectural that Old Icelandic had vowel harmony, but the possibility is supported by orthogonal evidence from insular runic inscriptions and certain manuscript material.

- statistical tendencies towards height harmonic distributions in certain Icelandic manuscripts have been interpreted as post-harmony decay remnants in Old Icelandic (Flom 1934a), and
even as far west as Greenland there are indications of height harmony; such as in the first of the Garðar stones (GR 1, ca. 13th–14th century) or the Kingittorsuaq stone (GR 1, ca. 1200/1250 or later) in Fig. 4 which contrast non-high gleðe gleðe ‘gladness’ or the name baanne Bjarne versus huilir hvílir ‘rests’ or fyrir fyrir ‘before’.

**Attested harmony decay:** Harmony decay has occurred in Nordic languages and is potentially captured in the Norse corpus.

**Philological descriptions**

Harmony is found in the earliest writing on parchment (c. mid-12th century)

- decaying gradually over the course of the late 13th and 14th centuries (Flom 1934b, Seip 1955, Hødnebø 1977, Hagland 1978)

As summarised by Hødnebø (1977):


This [vowel harmony] system can be seen from the oldest writings and up to a good ways into the 1300s as a kind of norm. Towards the end of the century, there is a steady decline with ever-increasing exceptions to the rule. (Hødnebø 1977: 379)

**Old Norwegian corpus summary**

**Old Norwegian philological material:**

- provides a sizeable corpus of manuscripts, charters, and runic inscriptions
- covering pre-, transitional, and post-decay stages of vowel harmony in the language
  
  making Old Norwegian a typologically highly significant specimen

**2.2 Methods and corpus**

**Methodological challenges**

Hødnebø’s generalisation is statistical

- we need to be able to quantify Old Norwegian harmony patterns

**Problem:** Old Norwegian corpus is largely inaccessible

- and poses significant philological problems
Digital corpora
Medieval Nordic Text Archive (MENOTA): https://menota.org/forside.xhtml

- an increasing, digitised sample of Old Norwegian manuscript material
  - many of which are lexically and morphologically tagged

<table>
<thead>
<tr>
<th>Abbr.</th>
<th>Signature</th>
<th>MS or work title</th>
<th>Date</th>
<th>Provenance</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>AM₁₄₃</td>
<td>AM 143 fol.</td>
<td>King’s Mirror</td>
<td>c 1275</td>
<td>Bergen</td>
<td>63900</td>
</tr>
<tr>
<td>DG₆₁₇₆₁₁₁</td>
<td>De la Gardie 4–7, fols. 1746–59v</td>
<td>Strenglekar–band 1</td>
<td>c 1270</td>
<td>Bergen</td>
<td>12913</td>
</tr>
<tr>
<td>DG₆₁₇₆₁₁₁</td>
<td>De la Gardie 4–7, fols. 30r–43v</td>
<td>Strenglekar–band 2</td>
<td>c 1270</td>
<td>Bergen</td>
<td>18640</td>
</tr>
<tr>
<td>DG₆₁₇₆₁₁₁</td>
<td>De la Gardie 8 fol, fols. 70r–10v</td>
<td>Legendary saga of St. Olaf</td>
<td>c 1223–50</td>
<td>Trondheim</td>
<td>40142</td>
</tr>
<tr>
<td>H₄₁₁</td>
<td>Holm perg 4 fol, fols. 15r–14v</td>
<td>Þiðriks saga af Bern</td>
<td>c 1277–1300</td>
<td>Uncertain</td>
<td>19603</td>
</tr>
<tr>
<td>H₄₁₁</td>
<td>Holm perg 4 fol, fols. 35r–44v</td>
<td>Þiðriks saga af Bern</td>
<td>c 1277–1300</td>
<td>Uncertain</td>
<td>64582</td>
</tr>
<tr>
<td>H₆</td>
<td>Holm perg 6 fol</td>
<td>Saga of Barlaam and Josaphat</td>
<td>c 1275</td>
<td>Eastern</td>
<td>76411</td>
</tr>
<tr>
<td>H₃₄</td>
<td>Holm perg 24 41v</td>
<td>Bøjarlǫg ok Farmannalǫg Magnúss Hákonarson</td>
<td>c 1273–1300</td>
<td>Bergen</td>
<td>11283</td>
</tr>
<tr>
<td>H₇</td>
<td>Holm perg 17 41v</td>
<td>Saga of Archbishop Thómas</td>
<td>c 1530</td>
<td>Uncertain</td>
<td>59884</td>
</tr>
<tr>
<td>NRA₆₁</td>
<td>NRA 61</td>
<td>A fragment of Karlamagnús saga</td>
<td>c 1250</td>
<td>Uncertain</td>
<td>983</td>
</tr>
</tbody>
</table>

Table 1: The Old Norwegian MENOTA corpus

Structure of MENOTA transcriptions

- MENOTA transcription of <höffgingi> ‘chieftain’-DAT.MSG.-DEF. (Holm perg 6 fol.)

  <w xml:id='w034581' lemma='höffgingi' me:msa='xNC gM nS cD sD'>
  <me:dipl>höffgingi</me:dipl>
  </w>

  With the help of Pavel Iosad (Edinburgh), we have written scripts, collecting:
  - word IDs
  - vowel patterns
  - morphological annotations
  - lemmas
  - etc.

A vowel harmony database
Using the data from the MENOTA corpus,

  - vowels are organised into pairwise sequences and evaluated for height agreement (2)
Harmonic span | $V_1$ | $V_2$ | $V_{1\_high}$ | $V_{2\_high}$ | $V_{H}$
---|---|---|---|---|---
{hoðîing|1|ianom} | $<o>$ | $<i>$ | False | True | False
{hoðîing|2|nom} | $<i>$ | $<a>$ | True | False | False
{hoðîing|3|ianom} | $<a>$ | $<o>$ | False | False | True

Table 2: Division into pairwise harmonic spans

Controlling for variation
An orthographic database like Tab. 2 is useful, but it needs to recover orthographic variation

* e.g. spelling variation for [o, ø, a] – <o, a>
  - more vowel phonemes than graphemes
  - [ø] generally has no unique letter in Norwegian writing
  * e.g. <hoðîingia> and <haðîingia> for normalised hoðîingja

Grapho-phonology
We need a way to triangulate between distinct etymological, phonological, and orthographic values for each given segment

• e.g. <hoðîingia> = $<ø-i> = [ø–i]$
• e.g. <haðîingia> = $<a-i> = [ø–i]$

Phonological/etymological annotations
Etymological annotations based on Holthausen (1948)

• encoded for the 600 most common lexemes in the corpus (220,418 words)

(7) Example annotation of root-initial vowels in *standa* 'stand' word forms
(8) Etymological vs. graphic representations of the *standa* 'stand' finite verb paradigm

<table>
<thead>
<tr>
<th></th>
<th>Present indicative</th>
<th>Present subjunctive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Singular</td>
<td>Plural</td>
</tr>
<tr>
<td>1.</td>
<td><em>stɛnd</em></td>
<td><em>stɛndum</em></td>
</tr>
<tr>
<td>2.</td>
<td><em>stɛndr</em></td>
<td><em>stɛndet</em></td>
</tr>
<tr>
<td>3.</td>
<td><em>stɛndr</em></td>
<td><em>standa</em></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Preterite indicative</th>
<th>Preterite subjunctive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Singular</td>
<td>Plural</td>
</tr>
<tr>
<td>1.</td>
<td><em>stoːð</em></td>
<td><em>stoːðom</em></td>
</tr>
<tr>
<td>2.</td>
<td><em>stoːtt</em></td>
<td><em>stoːðoð</em></td>
</tr>
<tr>
<td>3.</td>
<td><em>stoːð</em></td>
<td><em>stoːðo</em></td>
</tr>
</tbody>
</table>

**Sample annotated data**
An abbreviated example of the kind of vocalic data included in this database are provided in Table 3

### 2.3 Sanity check

#### 2.3.1 Vowel and word length frequencies

The manuscripts display good agreement on word length and vowel height class frequencies

![Graphs](image.png)

(a) Vowel height class frequencies  
(b) Word length frequencies

**Figure 6:** Vowel and word length frequencies by manuscript

**Some basic stats:**

- mean syllable length of 1.56
<table>
<thead>
<tr>
<th>id</th>
<th>dipl</th>
<th>expanded</th>
<th>lemma</th>
<th>seq_no</th>
<th>vi</th>
<th>v2</th>
<th>etym1</th>
<th>etym2</th>
<th>vi_high</th>
<th>v2_high</th>
<th>VH</th>
</tr>
</thead>
<tbody>
<tr>
<td>11857</td>
<td>stendr</td>
<td>stendr</td>
<td>standa</td>
<td>1</td>
<td>e</td>
<td>NA</td>
<td>Ε</td>
<td>NA</td>
<td>F</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>30331</td>
<td>støðe</td>
<td>støðe</td>
<td>standa</td>
<td>1</td>
<td>0</td>
<td>e</td>
<td>ο:</td>
<td>e</td>
<td>F</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>48957</td>
<td>stœðe</td>
<td>stœðe</td>
<td>standa</td>
<td>1</td>
<td>œe</td>
<td>e</td>
<td>ο:</td>
<td>e</td>
<td>F</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>65048</td>
<td>hofðinganom</td>
<td>hofðinganom</td>
<td>hofðingi</td>
<td>1</td>
<td>0</td>
<td>i</td>
<td>ɔ</td>
<td>i</td>
<td>F</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td>65048</td>
<td>hofðinganom</td>
<td>hofðinganom</td>
<td>hofðingi</td>
<td>2</td>
<td>ɪ</td>
<td>a</td>
<td>ɻ</td>
<td>a</td>
<td>T</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>65048</td>
<td>hofðinganom</td>
<td>hofðinganom</td>
<td>hofðingi</td>
<td>3</td>
<td>ɑ</td>
<td>o</td>
<td>a</td>
<td>o</td>
<td>F</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>34024</td>
<td>kkunar</td>
<td>kirkjunnar</td>
<td>kirkia</td>
<td>1</td>
<td>u</td>
<td>a</td>
<td>u</td>
<td>a</td>
<td>T</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>18773</td>
<td>giæva</td>
<td>giæva</td>
<td>giof</td>
<td>1</td>
<td>æ</td>
<td>a</td>
<td>æ</td>
<td>a</td>
<td>F</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>43804</td>
<td>giava</td>
<td>giava</td>
<td>giof</td>
<td>1</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>F</td>
<td>F</td>
<td>T</td>
</tr>
</tbody>
</table>

Table 3: Examples from the etymologically annotated data-frame (abbreviated)
• around 45% of words are polysyllabic
• median word length in writing is 4 letters
• average proportion of vowels to word length in writing is approximately 43% (1.64/4.02)

Sanity check conclusions
• vocalic and word length frequencies are uniform across the corpus
  ☞ coherent and consistent data

3 Visualising vowel harmony decay
With this detailed and richly annotated database, our goal is to:
• visualise and compare harmony frequencies across corpora
• tracking vowel harmony decay in Old Norwegian

3.1 PhonMatrix visualisations
PhonMatrix visualisations:
• coloured vowel association matrices
• provide easy visual discovery of harmony patterns
• developed by Mayer et al. (2010) and Mayer & Rohrdantz (2013)
  – accessible at http://phonmatrix.herokuapp.com/

PhonMatrix visualisations provide coloured matrices reflecting vowel co-occurrence frequencies, as in Fig. 7

PhonMatrix visualisations
PhonMatrix takes as an input a $V_1$–$V_2$ vowel matrix
• each vowel pair is assigned some association measure based on their frequency of occurrence
  – e.g. using the phi coefficient

The phi coefficient is a normalised measure of association based on the $\chi^2$ coefficient
• defined as the square root of the ratio of $\chi^2$ to the sample size
  – i.e. $\phi = \sqrt{\frac{\chi^2}{n}}$
Figure 7: PhonMatrix visualisations of pre-decay Old Norwegian harmony in H6

A practical illustration of how this is calculated is provided below using the crosstabulation in (9)

\[ \phi = \frac{vz - xy}{\sqrt{ab cd}} \]

(9) [a...e] contingency table

<table>
<thead>
<tr>
<th></th>
<th>[e]</th>
<th>not-e</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>[a]</td>
<td>v</td>
<td>x</td>
<td>a</td>
</tr>
<tr>
<td>not-a</td>
<td>y</td>
<td>z</td>
<td>b</td>
</tr>
<tr>
<td>Total</td>
<td>c</td>
<td>d</td>
<td></td>
</tr>
</tbody>
</table>

The phi coefficient ranges from -1 to 1

- PhonMatrix visualisation maps the phi values to a bipolar colour scale (from red to blue)
- the darkness of the colour provides a visual indicator of the strength of each V₁–V₂ association
  - positive associations are blue
  - negative associations are red

Some caveats:

- the PhonMatrix platform currently requires each segment to be monographic
  - i.e. aː and au are currently not permitted
- some vowels such as short [ø] occur too infrequently to provide reliable results – not included
Comparing harmony frequencies across the corpus

Figure 8: Sample of PhonMatrix visualisations of 13th-century Old Norwegian harmony decay

Fig. 8 illustrates the harmony patterns across the corpus in historically harmonising contexts
- including root-initial vowel sequences with potential harmony triggers and potential harmony targets
- excluding compounds, neutral segments, and other non-harmonising contexts

For clarity’s sake:
- I have added reference lines to Fig. 8, dividing high and non-high vowels
- manuscripts are ordered from highest average harmony levels to least
  - illustrating the range of harmony and harmony decay in the corpus

Summary of results:
- in pre-harmony decay manuscripts (H6/DG8)
  - $V_2$-[e, o] vowels (the $e/o$ columns) strongly correlate with non-high vowels [a, á, á, ø, é, é, o]
  - $V_2$-[i, u] vowels (the $i/u$ columns) pattern with high $V_1$-vowels [i, u, y, i, ú, ý]
    * resulting in the stark asymmetric distribution of blue/red [+]/[−] cells between high/non-high vowels

From left to right, this pattern is less and less discernable as the effect of harmony decay increases to completion in AM243/H17.
3.2 Distinguishing decay stages

![Graph showing mean vowel harmony levels by manuscript height class in pairwise sequences]

**Figure 9:** Mean harmony levels by manuscript height class in pairwise sequences

**Taking a broader look:**

- Fig. 9 illustrates mean height harmony percentages in historically harmonising $V_1$-$V_2$ sequences by $V_1$-height class
  - high vowels and diphthongs trigger high harmony
  - mid vowels and low vowels trigger non-high harmony

**Broad generalisations**

- lower mean vowel harmony (the reference line) is correlated with increasing dispersion
  - demonstrating that harmony decay is present in the corpus

**Pre-decay, robust harmony systems**

- manuscripts on the left (DG8–H4_h1)
- height correspondence is under tight control
  - high harmony and low variance

**Transitional systems**

- DG4_7 manuscripts
• probabilistic harmony
  – lower harmony but still low variance

Post-decay, non-harmonic systems
• manuscripts on the right (AM243–H17)
• no discernable height correspondence
  – low harmony and high variance
  – high or non-high height classes below 50% harmony threshold
  * reflects levelled, non-alternating suffixes

Coherent directions of change
The final stages of harmony decay result in levelled [-e, -o] or [-i, -u] suffixes
• both outcomes are documented in this corpus and found among modern Nordic languages, as illustrated in (10)

(10) Post-harmony inflectional systems

(a) AM243 – [-i] / [-um]

(hús-i) house-DAT.SG.  
(hús-um) house-DAT.PL.

ljós-i light-DAT.SG.  
ljós-um light-DAT.PL.

(b) H17 – [-e] / [-om]

(hús-e) house-DAT.SG.  
(hús-om) house-DAT.PL.

ljós-e light-DAT.SG.  
ljós-om light-DAT.PL.

(c) Mod.Icelandic – [-i] / [-um]

(hús-i) house-DAT.SG.  
(hús-um) house-DAT.PL.

ljós-i light-DAT.SG.  
ljós-um light-DAT.PL.

(d) Mod.Norw. (eastern) – [-e] / [-om]

(hús-e) house-DAT.SG.  
(hús-om) house-DAT.PL.

dal-e valley-DAT.SG.  
dal-om valley-DAT.PL.

4 Conclusions

Vowel harmony decay: rarely attested and poorly understood sound change
• Old Norwegian provides us with rare and typologically significant insights
  – but comes with significant philological and historical phonological challenges
Novel corpus linguistic methods:

- automated data collection using MENOTA transcriptions
- automated clean-up and linguistic annotation
- allows for easy stats/analysis

Generalisations:

1. Reliable phonological data despite philological challenges
   - consistent patterns and coherent directions of change
2. Documented vowel harmony change
   - clear spectrum of decaying harmony patterns
3. Distinct stages of decay
   - pre-decay: DG8, H6, H34, & H4_h1 – (high harmony - low variance)
   - transitional: DG4_7_h1/h2 – (medium harmony - medium variance)
   - post-decay: AM2,43, H4_h2, & H17 – (low harmony - high variance)

Limitations and speculations

Explanatory model:

- proof of concept
- only 10 scribes
  - far off from the kind of statistical power we’d need to explain why harmony decay has occurred
  
  ☹️ but we’ll get there :-) 

Aims and directions for future research

Decay diagnostics:

- aids for identifying changing harmony systems
  - e.g. Crimean Tatar (Turkic; Kavitskaya 2013); Kazakh (Turkic; McCollum 2015); Itelmen (Chukotko-Kamchatkan; Bobaljik 2018)

Historical dialectology:

- Old Norwegian harmony decay provides rich variation
  - can be used to better map geographic and chronological variation

Broader philological applications:

- These corpus methods allow for broader linguistic and philological study
  - e.g. sound mergers/splits, umlaut processes, scribal hand identification, etc.
References


Flom, George T. 1934a. Did Old Icelandic have vowel harmony? In Studia Germanica tillägnade Ernst Albin Kock, 45–58. Lund: Gleerup.


